

Genetic and environmental influences on psychological distress in the population: General Health Questionnaire analyses in UK twins

F. V. RIJSDIJK,¹ H. SNIEDER, J. ORMEL, P. SHAM, D. P. GOLDBERG
AND T. D. SPECTOR

From the Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry and Twin Research and Genetic Epidemiology Unit, St Thomas' Hospital, London; Georgia Prevention Institute, Medical College of Georgia, Augusta, GA, USA; and Department of Social Psychiatry, University of Groningen, The Netherlands

ABSTRACT

Background. The General Health Questionnaire (GHQ) is the most popular screening instrument for detecting psychiatric disorders in community samples. Using longitudinal data of a large sample of UK twin pairs, we explored (*i*) heritabilities of the four scales and the total score; (*ii*) the genetic stability over time; and (*iii*) the existence of differential heritable influences at the high (ill) and low (healthy) tail of the distribution.

Method. At baseline we assessed the GHQ in 627 MZ and 1323 DZ female pairs and at a second occasion (3.5 years later) for a small subsample (90 MZ and 270 DZ pairs). Liability threshold models and raw ordinal maximum likelihood were used to estimate twin correlations and to fit longitudinal genetic models. We estimated extreme group heritabilities of the GHQ distribution by using a model-fitting implementation of the DeFries–Fulker regression method for selected twin data.

Results. Heritabilities for Somatic Symptoms, Anxiety, Social Dysfunction, Depression and total score were 0.37, 0.40, 0.20, 0.42 and 0.44, respectively. The contribution of shared genetic factors to the correlations between time points is substantial for the total score (73%). Group heritabilities of 0.48 and 0.43 were estimated for the top and bottom 10% of the total GHQ score distribution, respectively.

Conclusion. The overall heritability of the GHQ as a measure of psychosocial distress was substantial (44%), with all scales having significant additive genetic influences that persisted across time periods. Extreme group analyses suggest that the genetic control of resilience is as important as the genetic control of vulnerability.

INTRODUCTION

In this paper genetic model-fitting results of the scaled General Health Questionnaire, measured longitudinally in a large sample of UK twin pairs, are presented. The General Health Questionnaire (GHQ) (Goldberg, 1972) is the most widely used screening instrument for detecting

psychiatric disorders in community and non-psychiatric clinical settings (Goldberg & Hillier, 1979; Goldberg & Williams, 1988). The GHQ items were chosen to differentiate psychiatric cases from non-cases. The content of the items refers to psychosocial distress and dysfunction. An individual's GHQ score gives a rough indication of the severity of psychological distress at the time the questionnaire is completed. This means that increasing total scores are associated

¹ Address for correspondence: Dr F. V. Rijdsdijk, SGDP Centre, Institute of Psychiatry, PO Box 80, London, SE5 8AF.

with increasing probability of being 'diagnosed' by a psychiatrist (Goldberg *et al.* 1997; Furukawa & Goldberg, 1999). The GHQ concentrates its items at the less differentiated level of emotional distress and does not aim to detect functional psychoses.

This is the first study to investigate the GHQ in a genetically sensitive design in a large sample of twins. What would be predicted about the genetic architecture? Since the GHQ-28 is a state measure (questioning 'the last couple of weeks') focusing on disruptions in normal functioning rather than on life-long traits or long-standing conditions, one would not expect to find high heritable influences determining individual differences in the scores. Heritability estimates should not be higher than those for trait anxiety or 1-year prevalence of major depression, which are in the range 30–45% (Kendler *et al.* 1994, 1995; Roy *et al.* 1995).

The longitudinal nature of our data enables the investigation of the genetic structure of the changes in psychiatric disturbance that occur with time (chronicity). Because the GHQ is a state measure, modest heritabilities and modest correlations between time point 1 and 2 are expected. Genetic factors are not expected to be the main cause of the phenotypic longitudinal correlation. We also tested the existence of differential heritable influences at the high (ill) and low (healthy) tail of the GHQ distribution.

METHOD

The study cohort was comprised of 749 monozygotic (MZ) and 1651 dizygotic (DZ) female twin pairs aged 18–79 years from the St Thomas' UK Adult Twin Registry (Andrew *et al.* 2001). Twins were ascertained from the general population through a UK nationwide media campaign conducted between 1992 and 1996 (for details see Spector *et al.* 1996). The twins were recruited to participate in a variety of studies without any exclusives or details of particular diseases or hypotheses (Hammond *et al.* 2000; Bataille *et al.* 2000). The majority of the volunteer sample consisted of healthy postmenopausal women who were invited to the Twin Research Unit at St Thomas' Hospital for a full medical examination. A higher proportion of DZ compared to MZ twins have been called for interview, as a major goal of subject

recruitment was to obtain genotype data for linkage studies, and MZ twins are not informative for linkage. All subjects completed a nurse-administered questionnaire, relating to their twin history, medical history and to environmental and life-style variables that could be potential confounders of the physical phenotypes under study. Among these variables were social class (as measured by present occupation of self and partner) and GHQ scores. Zygosity was determined by standard questionnaire method (Cederlöf *et al.* 1961), and DNA fingerprinting was used for confirmation.

The GHQ scores were obtained on attending St Thomas' Hospital. At the baseline visit, full-pair GHQ data were available for 627 MZ and 1323 DZ pairs (mean age, 47.7; s.d., 12.4). For a small group of early twins (90 MZ and 270 DZ pairs) who were invited back for a fuller examination, GHQ data was assessed on a second occasion. The mean time-lag between the first and second visit is 3.5 years (min = 0.9 years, max = 6.5 years). For discussing heritability estimates of the scale and total scores we will focus on the large baseline sample.

The General Health Questionnaire (GHQ)

The GHQ score was determined using the 28-item scaled version (GHQ-28). The scaled GHQ is derived by factor analysis and consists of four subscales of seven items each: Somatic Symptoms (factor A), Anxiety and Insomnia (factor B), Social Dysfunction (factor C) and Severe Depression (factor D). The questions ask subjects to compare their states in the past few weeks to their usual state, and measured the extent to which there was a discrepancy between the two. Responses were scored using a Likert scale (0–3) and the items are summed to generate scale scores assuming that they represent the scale to the same degree (Cronbach's alpha for scale A = 0.83, B = 0.88, C = 0.80 and D = 0.91). To separate cases *versus* normals in order to establish prevalence of illness in a population and to detect cases of 'hidden psychiatric illness' in general medical clinics, a threshold can be applied to the total score. For genetic model fitting, to estimate heritability components, we used the full distribution of the summed item score for each scale as well as the overall total score. To account for the wide age range, scales and the total score were age-regressed.

Statistics

Genetic model-fitting

Genetic model-fitting is used to investigate the relative contribution of genetic and environmental influences to individual differences in the scale scores and total GHQ score. The phenotypic variance is decomposed into: additive genetic variation, A, (i.e. the sum of the average effects of the individual alleles at all loci); shared environmental variation (C); and a unique environmental component (E) that is not shared by family members. No non-additive genetic influences (interactions between alleles at the same locus, dominance genetic variation) or on different loci (epistasis) were considered. From biometrical genetic theory we are able to write structural equations relating observed traits of twins to their underlying genotypes and environments. The relative magnitude and importance of these factors (the model parameters) can be inferred by fitting the observed correlations between twins to the predicted correlations according to the hypothesized model (ACE, AE, CE or E). A structural equation modelling program is used to estimate model parameters by minimizing a goodness-of-fit statistic (χ^2) between observed and predicted covariance matrices. Data can be entered as summary statistics (e.g. covariance matrices) or raw data. Raw data analysis allows greater flexibility, because it automatically handles many missing data problems.

Because all log-transformed age-regressed scale scores (except the Severe Depression scale) showed reasonable normality, raw maximum likelihood model fitting analyses were conducted on transformed continuous variables, using the program Mx (Neale, 1999). When analysing models with raw data, minus twice the log-likelihood ($-2 \cdot LL$) of the data for each observation is calculated. This implies that there is no overall measure of fit, but there are relative measures of fit, since differences in fit function between submodels are distributed as χ^2 . The goodness-of-fit of e.g. the full ACE model is measured relative to a perfect fitting (saturated) model. In a saturated model, the maximum number of parameters is estimated to describe the correlational structure between variables. For example, for the longitudinal correlation matrix in one group (MZ or DZ) that would be

10 parameters: four variances (of twin 1 and twin 2 at time 1 and 2); two within-twin cross-time correlations; two cross-twin within-time correlations; and two cross-twin cross-time correlations.

A non-significant χ^2 value suggests that the model is consistent with the data, whereas a significant χ^2 value ($P < 0.05$) suggests that the model poorly fits the data and can be rejected. Goodness-of-fit of alternative, nested models were evaluated by changes in χ^2 . For example, the fit of a reduced model (e.g. AE) will be better (i.e. the dropped parameter C will be non-significant) if the difference in χ^2 for one degree-of-freedom does not exceed the critical value (at the 0.05 level) of 3.84. Information about the precision of parameter estimates (and their explained variance) was obtained by likelihood-based confidence intervals (CIs) rather than standard errors (Neale & Miller, 1997).

Raw ordinal analysis

Since the Severe Depression scale was highly skewed, model fitting analyses were conducted on raw ordinalized data. Mx provides a method for analysing categorized data by using raw ordinal maximum likelihood estimation. Essentially, the model predicts proportions of twin pairs that should exist for the various possible patterns of responses from two twins, assuming a multivariate normal distribution. Observed frequencies in each category are translated in proportions under an assumed normal distribution by estimating associated thresholds (z values). Maximum likelihood for ordinal data provides better correlation estimates and better-fitting models for non-normally distributed data. In addition, the raw data approach facilitates inclusion of subjects with incomplete data. The age-regressed Severe Depression scores were ordinalized in four categories (three thresholds). Because variances cannot be estimated when thresholds are estimated, they are constrained to unity, and the number of parameters and observed statistics per zygosity group reduces to six in the saturated model.

A test of bivariate normality was performed on the categorised twin 1 and twin 2 Severe Depression scores. This test assesses the fit of a simple correlational model to the MZ and DZ, 4×4 contingency tables, in which the correlation and one set of thresholds were specified as free

parameters. Since the fit of this model is only marginally significant for both MZ and DZ pairs ($\chi^2_{(11)} = 20.8$, $P = 0.04$, $\chi^2_{(11)} = 20.5$, $P = 0.04$, respectively) and categorizing the scale in fewer classes (three or two) would mean considerable loss of valuable distributional information, we decided to perform analyses on the four-cotomous data.

Longitudinal genetic analyses

The difference in cross-twin within-time correlations between MZ and DZ pairs give an indication of the sources of variance in the model: DZ correlations that are approximately half that of the MZ correlations are an indication that heritable influences play a role. Similarly, the difference in cross-twin cross-time correlations between MZ and DZ pairs are indicative of the sources of variance that determine the correlation over time.

The genetic structure of the time 1 and time 2 data (and the correlation over time) was analysed in a bivariate (two-time point) genetic model. In the bivariate model the A, C and E matrix are specified in a triangular (Cholesky) decomposition. This decomposition implies as many (Genetic and Environmental) latent factors as variables. In, for example, the genetic matrix, factor A_1 influences the phenotype at time 1 (a_c) and time 2 (a'_c), whereas factor A_2 only influences the phenotype at time 2 (a_s). The heritability, c^2 and e^2 for each time point can be estimated, as well as the genetic, common environmental and unique environmental correlations between time points by standardizing the A, C and E matrix, respectively.

It is possible to partition the phenotypic correlation between time points in parts determined by common genes, common shared and unique environmental effects. The genetic contribution to the phenotypic correlation is calculated by $h_1 * r_g * h_2$, where r_g is the genetic correlation, and h_1 and h_2 are standardized additive genetic path estimates for time 1 and time 2, respectively (Plomin *et al.* 2001).

Extreme group heritabilities

We estimated the group heritability of the top and bottom 10% of the GHQ total score distribution. We used a model-fitting implementation of the DeFries-Fulker regression method for selected twin data (Purcell & Sham, 2003),

Table 1. Means (s.d.) of individual GHQ items, scale scores and total score of whole sample for time point 1 and 2

GHQ	Time 1 Mean (s.d.)	Time 2 Mean (s.d.)
A: Somatic Symptoms	3.87 (2.86) N=3969	4.04 (3.03) N=743
B: Anxiety	3.40 (2.84) N=4004	3.49 (2.86) N=741
C: Social Dysfunction	7.09 (2.22) N=3983	7.17 (2.15) N=747
D: Severe Depression	0.89 (2.03) N=3993	0.82 (1.90) N=750
Total Score	15.16 (7.61) N=3880	15.43 (7.71) N=724

where an individual's expected score is modelled as a function of their co-twin's proband status. This method allows the estimation of likelihood-based confidence intervals around the estimates.

RESULTS

Characteristics of the twin sample

Sociodemographic characteristics of the twin sample show close agreement with non-twin singleton samples (see for a detailed comparison Andrew *et al.* 2001). At the second sampling frame the twins were significantly older, 71% of the sample was older than 50, as compared to 47% at the first visit. As a result, the proportion of retired individuals at visit 2 is significantly higher than visit 1. The means and standard deviations of the GHQ factor scores and total score of both time points are listed in Table 1. The bias of older subjects at time 2 is not reflected in obvious differences in means between time points.

Heterogeneity of means/thresholds

The means of the log-transformed Somatic Symptoms, Anxiety and Social Dysfunction scores, could be constrained equal across twin 1 and twin 2 without significant decline in fit ($\Delta\chi^2_{(4)} = 36$, $P = 0.46$; 0.6 , $P = 0.96$; and 2.9 , $P = 0.57$, respectively). However, means were significantly different between MZ and DZ twin pairs for these scales at time 1 only ($\Delta\chi^2_{(1)} = 8.3$, $P = 0.004$; 7.1 , $P = 0.008$; and 9 , $P = 0.003$, respectively), with DZ pairs having higher means. Similar differences were observed for the means of the log-transformed total score at time 1

Table 2. Maximum likelihood estimation of twin-correlations (and 95% CI) for the GHQ scales at time 1 and 2. MZ correlations below, DZ correlations above diagonal

Somatic Symptoms	A1twin1	A2twin1	A1twin2	A2twin2
A1twin1	1	0.44	0.19 (0.14-0.24)	0.18 (0.07-0.29)
A2twin1	0.45	1	0.23 (0.13-0.33)	0.18 (0.07-0.29)
A1twin2	0.34 (0.28-0.41)	0.15 (-0.04-0.29)	1	0.52
A2twin2	0.22 (0.02-0.43)	0.40 (0.20-0.56)	0.29	1
Anxiety	B1twin1	B2twin1	B1twin2	B2twin2
B1twin1	1	0.44	0.20 (0.15-0.25)	0.15 (0.05-0.25)
B2twin1	0.55	1	0.25 (0.14-0.25)	0.35 (0.24-0.45)
B1twin2	0.40 (0.34-0.47)	0.30 (0.12-0.46)	1	0.51
B2twin2	0.34 (0.12-0.51)	0.40 (0.20-0.56)	0.46	1
Social Dysfunction	C1twin1	C2twin1	C1twin2	C2twin2
C1twin1	1	0.22	0.08 (0.03-0.14)	0.13 (0.02-0.25)
C2twin1	0.15	1	0.13 (0.01-0.24)	0.14 (0.02-0.25)
C1twin2	0.21 (0.14-0.29)	0.14 (-0.03-0.31)	1	0.21
C2twin2	0.07 (-0.12-0.26)	0.30 (0.1-0.48)	0.14	1
Depression	D1twin1	D2twin1	D1twin2	D2twin2
D1twin1	1	0.45	0.24 (0.18-0.30)	0.22 (0.13-0.34)
D2twin1	0.50	1	0.21 (0.08-0.33)	0.22 (0.09-0.35)
D1twin2	0.40 (0.31-0.47)	0.12 (-0.11-0.34)	1	0.58
D2twin2	0.31 (0.05-0.52)	0.33 (0.10-0.54)	0.28	1
Total Score	T1twin1	T2twin1	T1twin2	T2twin2
T1twin1	1	0.45	0.24 (0.19-0.29)	0.20 (0.09-0.30)
T2twin1	0.37	1	0.25 (0.14-0.35)	0.29 (0.19-0.41)
T1twin2	0.42 (0.35-0.48)	0.24 (0.06-0.41)	1	0.51
T2twin2	0.17 (-0.10-0.38)	0.44 (0.25-0.59)	0.33	1

Within-twin cross-time correlations (*italic*); cross-twin within-time correlations (**bold**); cross-twin cross-time correlations (*italic bold*). MZ correlations are below, DZ correlations above diagonal.

($\Delta\chi^2_{(1)} = 10.5$, $P = 0.001$). For the Depression scale, thresholds could be constrained to be the same across twins ($\Delta\chi^2_{(12)} = 9.2$, $P = 0.69$) but not across zygosity groups ($\Delta\chi^2_{(6)} = 23$, $P < 0.001$). These differences in means across zygosity groups are addressed in the discussion.

Longitudinal (bivariate) genetic model-fitting

Table 2 shows the maximum likelihood estimates of the within-time-point cross twins, the cross-time-point within twin, and the cross-time-point cross twins correlations (with 95% CIs). These correlations suggest additive genetic and shared environmental influences for explaining familiarity rather than dominance genetic effects. The genetic model fitting results are presented in Table 3. For all scale scores (A, B, C and D) and the total score common environmental influences could be dropped from the full ACE model without significant decline in fit ($\Delta\chi^2_{(3)} = 0.1, 4.4, 0.1, 2$ and 1.6 , respectively), whereas this was not the case for additive genetic effects. The model in which individual differences could be explained by unique environmental effects only (Model E) was rejected for all variables. In Table 4, heritability

estimates under the full ACE as well as the best-fitting models are given for time 1 and 2. At time 1, heritability estimates for Somatic Symptoms, Anxiety and Severe Depression were substantial: 37%, 40% and 42%, respectively. For Social Dysfunction, h^2 was estimated to be lower (20%). The total GHQ score showed the highest heritability: 44% at time 1, and 51% at time 2, with lower CIs of around 39% and upper CIs of around 60%.

Genetic stability

Indices of genetic and environmental influences determining chronicity/stability over time are given in Table 5. The genetic correlations between time points for all variables are high, ranging from 0.69 to 0.80. These high genetic correlations indicate that genes impacting at time 1 are likely to also affect time 2. The contribution of shared genetic factors to the phenotypic correlations between time 1 and time 2 are high for Somatic Symptoms, Anxiety, Depression and GHQ total score and estimated to be between 65% and 73%. This implies that environmental factors contribute less to the stability/chronicity over time for these variables

Table 3. Raw (Ordinal) model fitting results for longitudinal GHQ scale scores and GHQ total score

	Genetic models											
	ACE			AE			CE			E		
	-2LL	df	$\chi^2_{(11)}$	-2LL	df	$\chi^2_{(10)}$	-2LL	df	$\chi^2_{(10)}$	-2LL	df	$\Delta\chi^2_{(11)}$
Continuous												
A: Somatic Symptoms	27269.5	4672	27.7**	27269.6†	4675	27.8*	27288.8	4675	47***	27413.1	4678	171.3***
B: Anxiety	30981.4	4721	9.5	30985.8†	4724	13.9	31006.1	4724	34.2**	31185.8	4727	213.9***
C: Social Dysfunction	20774.5	4711	12.7	20774.6†	4714	12.8	20784.2	4714	22.4***	20892.3	4717	67.5***
Total Score	2641.8	4587	13.5	2643.4†	4590	15.1	2666.2	4790	37.9***	2867.6	4593	239.3***
Ordinal												
D: Severe Depression	12667.8	4682	11.1*	12669.8†	4685	13.1	12678.6	4685	21.9**	12817.8	4688	161.1***

For the ordinal analysis, the 5 df for the full bivariate ACE model is derived by: 12 (20 observed statistics minus 8 variance constraints) - 7 (9 estimated parameters minus 2 variance constraints). For the continuous analyses, the 11 df for the full bivariate ACE model is derived by: 20 observed statistics minus 9 estimated parameters; χ^2 values are derived by, -2LL of Genetic model minus -2LL of Saturated model.

† Best-fitting model.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; all other χ^2 comparisons, non-significant (> 0.05).

Table 4. Standardized estimates (with 95% CI) of full ACE and best-fitting models for the GHQ scales and total score at two time points

Scales	Time 1			Time 2		
	h^2	c^2	e^2	h^2	c^2	e^2
Somatic Symptoms	0.35 (0.18-0.42) 0.37 (0.35-0.42)	0.02 (0.00-0.13) —	0.63 (0.58-0.71) 0.63 (0.58-0.69)	0.37 (0.04-0.52) 0.39 (0.25-0.52)	0.02 (0.00-0.22) —	0.61 (0.48-0.78) 0.61 (0.48-0.75)
Anxiety	0.38 (0.25-0.45) 0.40 (0.35-0.46)	0.02 (0.00-0.11) —	0.60 (0.55-0.60) 0.60 (0.55-0.64)	0.18 (0.01-0.50) 0.50 (0.37-0.60)	0.24 (0.02-0.39) —	0.58 (0.44-0.70) 0.50 (0.40-0.63)
Social Dysfunction	0.20 (0.07-0.26) 0.20 (0.14-0.26)	0.00 (0.00-0.09) —	0.80 (0.74-0.86) 0.80 (0.74-0.86)	0.31 (0.01-0.46) 0.31 (0.15-0.46)	0.00 (0.00-0.20) —	0.69 (0.54-0.87) 0.69 (0.54-0.85)
Severe Depression	0.31 (0.12-0.46) 0.42 (0.36-0.48)	0.09 (0.00-0.23) —	0.60 (0.53-0.68) 0.58 (0.52-0.64)	0.26 (0.00-0.51) 0.39 (0.23-0.53)	0.10 (0.00-0.33) —	0.64 (0.48-0.83) 0.61 (0.47-0.77)
Total Score	0.38 (0.23-0.48) 0.44 (0.39-0.50)	0.05 (0.00-0.16) —	0.57 (0.51-0.63) 0.56 (0.50-0.61)	0.35 (0.00-0.60) 0.51 (0.38-0.62)	0.12 (0.00-0.36) —	0.53 (0.40-0.71) 0.49 (0.38-0.62)

h^2 , Heritability.

(only 27% for the GHQ total score). For Social Dysfunction, environmental and genetic influences contribute equally to the phenotypic correlation (50% each).

Extreme group heritabilities of the GHQ distribution

To test the hypothesis that heritable influences may exist at the low end as much as at the high end of the GHQ, influencing the tendency to have a persistently low score as well as a persistently high one, we performed extreme group heritability analyses (Purcell & Sham, 2003). Two groups were analysed: probands selected for having a score in the top 10%, and probands scoring in the lower 10% of the distribution.

For the top 10% selection (85 MZ and 257 DZ pairs), common environmental influences

were non-significant ($\Delta\chi^2_{(1)} = 0.1$, $P = 0.75$), and estimated group h^2 of the AE model was 0.48 (95% CI: 0.39-0.56). Analysis of the lower 10% group (99 MZ and 224 DZ pairs), showed a similar pattern. Common environmental influences were nonsignificant ($\Delta\chi^2_{(1)} = 0.1$, $P = 0.75$), and estimated group h^2 of the AE model was 0.43 (95% CI: 0.33-0.53).

DISCUSSION

This is the first study to investigate the genetic structure of the GHQ-28 in a twin sample measured at two time points. For none of the scales, common environmental influences were found to be significant. Additive genetic effects sufficiently explained family resemblance in all scale scores and the total GHQ score.

Table 5. Partitioning of longitudinal phenotypic correlations of GHQ scale scores and GHQ total score

Scale	Additive genetic paths		Non-shared environmental paths		r_g r_e		Contribution to phenotypic correlation by		% of phenotypic correlation accounted for by A
	Time 1	Time 2	Time 1	Time 2			A	E	
Som Symp	0.61	0.62	0.79	0.78	0.80	0.26	0.30	0.16	65
Anxiety	0.63	0.71	0.78	0.71	0.75	0.26	0.34	0.14	71
Social Dysf	0.45	0.56	0.89	0.83	0.75	0.26	0.19	0.19	50
Depression	0.64	0.62	0.76	0.78	0.80	0.29	0.32	0.17	65
Total Score	0.66	0.71	0.75	0.70	0.69	0.22	0.33	0.12	73

The phenotypic correlation is the sum of contributions by A and E; r_g = genetic correlation; r_e = correlation due to E.
Som Symp, Somatic Symptoms; Social Dysf, Social Dysfunction.

Heritability estimates were substantial, ranging from 20% (Social Dysfunction) to 44% (total score). For all but the Social Dysfunction scale scores and the total score, genetic influences, contribute more to the phenotypic correlation between time 1 and time 2 than do environmental influences.

How can we explain these substantial heritability estimates and the relatively high genetic correlations over time given that the GHQ-28 is a state rather than a trait measure? The heritability estimate for the Anxiety scale (40%) is in accordance with estimates reported by other studies (Roy *et al.* 1995). Heritability estimates, reported for depression, vary as a result of both the definition of depression used and the group in which it is measured. Severe depression in hospitalized selected cases show high heritability estimates of around 80% (McGuffin *et al.* 1996). Questionnaire data of the normal range of depressive symptoms in unselected twin samples show small genetic influences (15%) if symptoms were reported with regard to 'the last week' (Kendler *et al.* 1991) and moderate heritabilities (30–37%) if symptoms regarded 'past 30 days' (Kendler *et al.* 1994). Interview based DSM-III-R data showed heritability estimates for 1-year prevalence of major depression of 42% (Kendler *et al.* 1992, 1993a), whereas genetic influences for lifetime history of major depression was reported to be between 63–83% (Kendler *et al.* 1995).

Our heritability estimate for Severe Depression measured over the 'last couple of weeks' (42%) and the heritability of the total GHQ score (44%) are in accordance with reported heritabilities of 1-year prevalence of major depression (Kendler *et al.* 1992, 1993a),

and thus, higher than expected. We also observe that the test-retest correlations between visit 1 and 2 are substantial (between 0.7–0.8) especially given the long mean time lapse of 3.5 years. This stability is consistent with other personality-like traits (Duncan-Jones *et al.* 1990; Kendler *et al.* 1994). Kendler *et al.* (1994) suggested that, although self-report questionnaires are designed to measure 'state' depressive symptoms and anxiety, they might to a large extent reflect stable traits like neuroticism. An additional support for this explanation, given by our data, is the high genetic correlation between time points, indicating that genes impacting at time 1 are likely to also affect time 2. Secondly, genetic factors substantially contribute (73%) to the phenotypic correlations between time 1 and time 2 of the total GHQ score. This effect is more likely to occur for trait-like attributes i.e. 1-year prevalence of major depression for which genetic effects were found to be entirely stable in a one-year interval (Kendler *et al.* 1993a).

However, there are other possible mechanisms to explain substantial heritabilities for self-report questionnaire 'state' scores. First, questionnaire data may reflect the reporting rather than the actual occurrence of events/symptoms. When relying on these 'indirect' assessments we can not rule out the possibility that the results are due to genetic effects on 'reporting behaviour' or the tendency to 'complain' (Kendler & Karkowski, 1997). Another possible explanation for relatively high heritability estimates of depression is the 'genetic control of exposure to the environment' or genetic correlation. This means that genetic factors influencing the liability to e.g. major depression may in part exert their effect indirectly by altering an individual's

probability of exposure to depressogenic environments. The first indication for this effect is the fact that even 'environmental' variables like stressful life events show genetic influences between 20 and 40% in children, adolescents and adults (Plomin *et al.* 1990; Kendler *et al.* 1993*b*; Billig *et al.* 1996; Thapar & McGuffin, 1996) and the existence of a genetic correlation between the environmental measures and the trait. Kendler & Karkowski (1997) explored the relationship between genes for life events and depression: a high genetic risk for depression was associated with an increased risk of experiencing life events and 10–15% of the impact of genes on risk for major depression is mediated through stressful life events. When genetic correlation exists and is not controlled for, the effect will be estimated in the heritability component (Neale & Cardon, 1992).

We investigated the existence of differential heritable influences at the high and low extremes of the GHQ distribution. Group heritabilities for the top and bottom 10% tails of the distribution showed to be very similar (48% and 43%, respectively) with largely overlapping confidence intervals. This result is remarkable since this questionnaire was designed as a screening instrument to detect psychiatric cases in the community, and, therefore, discrimination is expected to be better at the top end of the scale. These results suggest that the genetic control of resilience seems to be as important as the genetic control of vulnerability.

The findings should be interpreted in the context of some potentially important methodological limitations. First, the results (heritability estimates and genetic stability over time) are applicable to (middle-aged) females, and may not generalize to other groups in the population. Secondly, this twin sample is from a volunteer register based on recruitment through advertisements. Volunteer bias for participants interested in medical research can not be totally ruled out. However, sociodemographics of this twin sample show close agreement with non-twin singleton samples (Andrew *et al.* 2001). Also, the effects of sex, age, and social class on the GHQ scores at time point 1 are in agreement with the effects found in the 'normal' population (Goldberg & Hillier, 1979). Thirdly, the longitudinal results should also be interpreted while noting that at the second visit the twins were

significantly older: 71% of the sample was older than 50, as compared to 47% at the first visit. However, this effect is not reflected in obvious differences in means between time points. Fourthly, at time point 1, DZ pairs have significantly higher means on all scales and the total GHQ score. A possible explanation for this, is that the effort of testing more DZ compared to MZ twins, may have led to ascertainment of more 'depressed', and ill-motivated DZ twins, who were only willing to be tested and interviewed after two or more invitations. It is not clear how the distortion of means might influence the DZ correlations and, therefore, the results of the genetic model fitting. The differences in means across zygosity groups were accounted for in all genetic analyses (i.e. were not constrained to be equal).

REFERENCES

- Andrew, T., Hart, D. J., Snieder, H., de Lange, M., Spector, T. D. & MacGregor, A. J. (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Research* 4, 464–477.
- Bataille, V., Snieder, H., MacGregor, A. J., Sasieni, P. & Spector, T. D. (2000). Genetics of risk factors for melanoma: an adult twin study of nevi and freckles. *Journal of the National Cancer Institute* 92, 457–463.
- Billig, J. P., Hershberger, S. L., Iacono, W. G. & McGue, M. (1996). Life events and personality in late adolescence: genetic and environmental relations. *Behavior Genetics* 26, 543–554.
- Cederlöf, R., Friberg, L., Jonsson, E. & Kaij, L. (1961). Studies on similarity diagnosis in twins by means of mailed questionnaire. *Acta Genetica* 338–362.
- Duncan-Jones, P., Fergusson, D. M., Ormel, J. & Horwood, L. J. (1990). *A Model of Stability and Change in Minor Psychiatric Symptoms: Results from Three Longitudinal Studies. Psychological Medicine, Monograph Supplement*, 18. Cambridge University Press: Cambridge.
- Furukawa, T. & Goldberg, D. P. (1999). Cultural invariance of likelihood ratios for the GHQ. *Lancet* 353, 561–562.
- Goldberg, D. P. (1972). *The Detection of Psychiatric Illness by Questionnaire. Maudsley Monograph*, 21. Oxford University Press: Oxford.
- Goldberg, D. P. & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine* 9, 139–145.
- Goldberg, D. P. & Williams, P. (1988). *A User's Guide to the General Health Questionnaire*. NFER-NELSON: Windsor.
- Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O. & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine* 27, 191–197.
- Hammond, C. J., Snieder, H., Spector, T. D. & Gilbert, C. E. (2000). Genetic and environmental factors in age-related nuclear cataracts in monozygotic and dizygotic twins. *New England Journal of Medicine* 342, 1786–1790.
- Kendler, K. S. & Karkowski-Shuman L. (1997). Stressful life events and genetic liability to major depression: genetic control of exposure to the environment. *Psychological Medicine* 27, 539–547.
- Kendler, K. S., Neale, M. C., Heath, A. C., Kessler, R. C. & Eaves, L. J. (1991). Life events and depressive symptoms: a twin study perspective. In *The New Genetics of Mental Illness* (ed. P. McGuffin

- and R. Murray), pp. 146–164. Butterworth-Heinemann Ltd: Oxford.
- Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. (1992). A population-based twin study of major depression in women. The impact of varying definitions of illness. *Archives of General Psychiatry* **49**, 257–266.
- Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. (1993a). A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry* **50**, 853–862.
- Kendler, K. S., Neale, M. C., Kessler, R., Heath, A. C. & Eaves, L. J. (1993b). A twin study of recent life events and difficulties. *Archives of General Psychiatry* **50**, 789–796.
- Kendler, K. S., Walters, E. E., Truett, K. R., Heath, A. C., Neale, M. C., Martin, N. G. & Eaves, L. J. (1994). Sources of individual differences in depressive symptoms: analysis of two samples of twins and their families. *American Journal of Psychiatry* **151**, 1605–1614.
- Kendler, K. S., Pederson, N. L., Neale, M. C. & Mathe, A. A. (1995). A pilot Swedish twin study of affective illness including hospital- and population-ascertained subsample: results of model fitting. *Behavior Genetics* **25**, 217–232.
- McGuffin, P., Katz, R., Watkins, S. & Rutherford, J. (1996). A hospital-based twin register of the heritability of DSM-IV unipolar depression. *Archives of General Psychiatry* **53**, 129–136.
- Neale, M. C. (1999). *Mx: Statistical Modeling*. Department of Psychiatry: Box 710 MCV, Richmond, VA 23298.
- Neale, M. C. & Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*. Kluwer Academic Publishers: Dordrecht.
- Neale, M. C. & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics* **27**, 113–120.
- Plomin, R., Lichtenstein, P., Pedersen, N. L., McClearn, G. E. & Nesselroade, J. R. (1990). Genetic influence on life events during the last half of the life span. *Psychology & Aging* **5**, 25–30.
- Plomin, R., DeFries, J. C., McClearn, G. E. & McGuffin, P. (2001). *Behavioral Genetics, 4th edn*. Worth Publishers: New York.
- Purcell, S. & Sham, P. C. (2003). A model-fitting implementation of the DeFries–Fulker model for selected twin data. (In the press).
- Roy, M. A., Neale, M. C., Pederson, N. L., Mathe, A. A. & Kendler, K. S. (1995). A twin study on generalized anxiety disorder and major depression. *Psychological Medicine* **25**, 1037–1049.
- Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J. & Hart, D. (1996). Genetic influences on osteoarthritis in women: a twin study. *British Medical Journal* **312**, 940–943.
- Thapar, A. & McGuffin, P. (1996). The genetic etiology of childhood depressive symptoms: a developmental perspective. *Development & Psychopathology* **8**, 751–760.